نوروزی، یعقوب؛ هماوندی، هدی (۱۳۹۴). بررسی مشکلات جستجو و بازیابی تصاویر در موتورهای کاوش برگزیده مبتنی بر ویژگی‌های نگارشی زبان فارسی. *پژوهش‌نامه کتابداری و اطلاع‌رسانی*، ۵ (۲)، ۲۲۲-۲۰۶.

بررسی مشکلات جستجو و بازیابی تصاویر در موتورهای کاوش برگزیده مبتنی بر ویژگی‌های نگارشی زبان فارسی

دکتر یعقوب نوروزی[۱]، هدی هماوندی[۲]

تاریخ دریافت: ۹۲/۱۰/۱۷ تاریخ پذیرش: ۹۳/۵/۱

چکیده

هدف: *پژوهش حاضر با هدف تعیین مشکلات جستجو و بازیابی تصاویر در موتورهای کاوش برگزیده مبتنی بر ویژگی‌های نگارشی زبان فارسی انجام شد.*

روش: *این پژوهش از نوع کاربردی است و برای پاسخ‌گویی به سؤالات پژوهش از روش ارزیابی با مشاهده مستقیم استفاده شد. جامعه آماری پژوهش شامل سه موتور کاوش گوگل، یاهو و بینگ است. برای گردآوری داده‌ها، از سیاهه محقق ساخته استفاده شد و تجزیه و تحلیل داده‌ها در دو سطح آمار توصیفی و استنباطی صورت گرفت.*

یافته‌ها: *نتایج پژوهش نشان داد که موتورهای کاوش گوگل، بینگ و یاهو بسیاری از ویژگی‌های نوشتاری و معنایی زبان فارسی را در هنگام جستجو و بازیابی تصاویر نادیده می‌گیرند. همچنین مشکلات مربوط به نگارش واژگان مشتق، مشتق- مرکب، انواع جمع‌های فارسی و مکسر عربی، همزه بدون کرسی و استفاده از زبان محاوره در بخش نوشتاری؛ و چند معنایی در بخش معنایی از مشکلات عمده زبان فارسی در جستجو و بازیابی تصاویر از این موتورهای کاوش به‌شمار می‌آیند. به‌علاوه، در بخش معنایی و نوشتاری، قابلیت‌های گوگل در انطباق با زبان فارسی نسبت به همتایانش بیشتر است.*

کلیدواژه‌ها: *بازیابی اطلاعات، بازیابی تصویر، موتور جستجو، شیوه نگارش، زبان فارسی*

۱. دانشیار گروه علم اطلاعات و دانش‌شناسی دانشگاه قم، ynorouzi@gmail.com

۲. دانشجوی دکتری علم اطلاعات و دانش شناسی، H.homavandi@gmail.com

مقدمه

تلاش همیشگی انسان‌ها برای ثبت و ضبط و گردآوری و اشاعه اطلاعات و دانسته‌های خود، موجب شد شاهد افزایش قابل‌ملاحظه انتشار و تولید روزافزون اطلاعات در تمامی زمینه‌ها به شکل چاپی و الکترونیکی باشیم (داورپناه، ۱۳۸۷). راه‌هایی نیز برای نشر این حجم انبوه اطلاعات به‌وجود آمده است که از جمله پرکاربردترین آنها شبکه گسترده جهانی وب است. از ابزارهایی که به‌منظور سازماندهی و تجمیع اطلاعات در فضای وب کاربرد زیادی دارند، موتورهای جستجو[1] هستند که تبدیل به یکی از سریع‌ترین و ساده‌ترین راه‌ها برای کاوش و یافتن اطلاعات در حوزه‌های مختلف و در قالب‌های گوناگون نظیر متن، صدا، تصویر و چندرسانه‌ای شده‌اند. در همین راستا تصاویر از اهمیت ویژه‌ای برخوردارند؛ چراکه در مواردی یک تصویر می‌تواند از صدها کلمه گویاتر باشد. لی‌یو و همکاران[2] (۲۰۰۷) معتقدند که همزمان با توسعه وب و دسترسی به انواع فناوری‌های عکس‌برداری مثل دوربین‌های دیجیتال و پویشگرهای[3] مخصوص تصاویر، اندازه مجموعه‌های تصاویر دیجیتال هم به‌سرعت رو به افزایش است. به همین دلیل نیاز به ابزارهایی کارآمد برای جستجو، مرور و بازیابی تصاویر با دامنه و موضوعات گوناگون افزایش یافته است. برای نیل به اهداف یاد شده سامانه‌های بازیابی تصاویر زیادی ایجاد شده‌اند که یکی از مهم‌ترین آنها موتورهای کاوش هستند؛ اما استفاده از این موتورهای کاوش با مسائل و مشکلاتی نیز روبرو است. نوتس[4](۱۹۹۷؛ نقل در نیازی، ۱۳۸۲) معتقد است، ردیابی سریع اطلاعات موردنیاز در اینترنت به‌صورت نیازی پیچیده درآمده است. از دلایل عمده این پیچیدگی می‌تواند گوناگونی کاربران، زبان‌ها، و فرهنگ‌های آنان باشد. باوجود اینکه موتورهای جستجوی متعددی برای تسهیل جستجو در محیط وب وجود دارند، به نظر می‌رسد که توجه آنها به زبان‌های غیرانگلیسی در مقایسه با انگلیسی کافی نیست (Lazarinis, A, 2007). آمارهای مربوط به سال ۲۰۱۵ در مورد استفاده از اینترنت براساس زبان حاکی از آن است که، حدود ۶۲/۴ درصد کاربران انگلیسی زبان و ۳۷/۶ درصد غیرانگلیسی زبان هستند (وب‌سایت آمارهای جهانی اینترنت، ۲۰۱۵). کاربران فارسی‌زبان از جمله ایران نیز در زمره این گروه هستند، چراکه براساس آخرین آمار اعلام شده در سال ۱۳۹۲ از سوی مرکز مدیریت توسعه ملی اینترنت بالغ بر ۶۱/۰۶ درصد جمعیت کشور به اینترنت متصل می‌شوند[5] که توجه

1. Search Engine
2. Liu & et al.
3. Scanner
4. Notess

۵. خبرگزاری مهر، ۱۸ مرداد ۱۳۹۲، جدیدترین آمار از ضریب نفوذ اینترنت در ایران، تاریخ دسترسی ۲۵ مرداد ۱۳۹۲، نشانی دسترسی http://www.mehrnews.com/detail/News/2093265.

به رفع نیازهای اطلاعاتی این گروه و تشخیص مشکلات زبانی و خطی آنها در استفاده از موتورهای کاوش امری اجتناب‌ناپذیر به نظر می‌رسد.

با توجه به موارد یاد شده، اهمیت استخراج اطلاعات از محیط وب و به‌خصوص موتورهای کاوش مشخص می‌شود و در این فرایند، زبان جستجو از مهم‌ترین و کلیدی‌ترین موارد است که حساسیت آن در رابطه با کاوش و بازیابی تصاویر دوچندان می‌شود چراکه «یکی از تفاوت‌های اساسی میان اطلاعات متنی و دیداری، ماهیت فرایند بازیابی آنها است. بازیابی اطلاعات متنی، مبتنی بر کشف شباهت‌های معنایی و نحوی بین موجودیت‌های متنی است. در حالی‌که بازیابی اطلاعات دیداری، مبتنی بر کشف شباهت‌های ادراکی و تداعی ذهنی است» (Zachary & et al, 2001). موتورهای جستجوی تصاویر، تصاویر را از طریق کلیدواژه‌ها، عنوان، برچسب‌ها، برچسب جایگزین[۱]، مشخصه‌های رنگ، بافت، حالت و شکل تصویر که به‌طور خودکار قابل استخراج از تصاویر هستند، بازیابی می‌کنند (Brown & Jeremy, 2007 نقل در نوروزی و ولایتی، ۱۳۸۹). حال آنکه بازیابی واژه‌ها در متون براساس بافت و زمینه کاربرد واژه در متن و جمله انجام می‌شود، اما در مورد تصاویر اگر عنوان متنی موجود نباشد اختصاص برچسب جهت شناسایی و نمایه‌سازی تصویر به‌وسیله موتورهای کاوش اجتناب‌ناپذیر است که این امر یعنی نحوه و شیوه نگارش برچسب اختصاصی نیز بر حساسیت و تفاوت‌های جستجو و بازیابی تصاویر می‌افزاید. به‌عنوان نمونه اختصاص برچسب «شیر» به تصویری از یک شیر آب در مرحله کاوش و بازیابی برای کاربرانی که در جستجوی تصاویری از شیر به‌عنوان ماده لبنی یا شیر به معنی حیوان هستند مسئله‌ساز خواهد بود. «بنابراین موتورهای کاوش تصاویر برای نمایه‌سازی به متن متکی هستند بدین معنی که کیفیت نتایج بازیابی شده آنها به کیفیت اطلاعات متنی اطراف یک تصویر و یا همراه با آن (مانند نام فایل، متن مجاور تصویر، عنوان صفحه و یا برچسب اچ تی ام آل) وابسته است» (TASI[۲], 2008). هر زبانی از جمله فارسی؛ دارای ویژگی‌ها و ظرایفی است که عدم توجه به آنها می‌تواند موجب ایجاد مشکلاتی در کاوش و بازیابی اطلاعات شود. از همین رو پژوهش حاضر تلاش دارد تا مشکلات جستجو و بازیابی تصاویر را به زبان فارسی از موتورهای کاوش

---

۱. در صفحه‌های وب هنگام ایجاد پیوند به یک تصویر در قسمت «برچسب جایگزین» عنوان تصویر وارد می‌شود تا هنگام نمایه‌سازی تصویر توسط موتورهای جستجو از اطلاعات برچسب جایگزین به‌عنوان کلیدواژه‌های مرتبط با آن تصویر استفاده شود و در صورت عدم‌بارگذاری یک تصویر متن برچسب جایگزین توسط جستجوگران قابل‌رؤیت باشد (نوروزی و ولایتی، ۱۳۸۹).

2 .TASI (Technical Advisory Service for Images)

برگزیده عمومی خارجی گوگل، بینگ و یاهو[1]، که براساس رتبه‌بندی‌ها، جزء پراستفاده‌ترین موتورهای جستجو با قابلیت پشتیبانی زبان فارسی و هم‌چنین قابلیت جستجو و بازیابی تصاویر هستند، بررسی نموده و ضمن شناسایی عمده‌ترین این مشکلات (معنایی و نوشتاری[2])، عملکرد آنها را در بازیابی تصاویر به زبان فارسی و براساس ویژگی‌های نوشتاری و معنایی[3] این زبان، با یکدیگر مقایسه نماید و کارآمدترین موتور کاوش جهت انطباق با زبان فارسی در بازیابی تصاویر را تعیین نماید تا به این ترتیب علاوه بر آزمودن مصادیق این قبیل مسائل، راه‌حل‌هایی را نیز جهت پیشگیری و رفع آنها مبتنی بر نتایج پژوهش ارائه کند.

ویژگی‌های نگارشی زبان فارسی

در رابطه با ویژگی‌های نگارشی زبان فارسی مرتضائی (۱۳۸۰) دسته‌بندی‌هایی را ارائه کرده است که شماری از آنها به انضمام سایر شاخصه‌های قابل‌تأمل از دیدگاه صاحب‌نظران حوزه زبان و ادبیات فارسی در ادامه آمده است.

الف- ویژگی‌های نوشتاری: این ویژگی‌ها به‌طور خلاصه شامل: گوناگونی در برگردان و ضبط واژگان بیگانه، عدم یکپارچگی در کاربرد کلمات دخیل از زبان‌های دیگر و معادل پیشنهاد شده از سوی فرهنگستان زبان و ادب فارسی، پیوسته‌نویسی و جدانویسی واژگان مشتق و مرکب و علائم جمع، تعدد علائم جمع (ها، ان، ات، ین، ون) و وجود جمع بی‌قاعده عربی در زبان فارسی، استفاده از تای منقوط، نحوه نگارش همزه میانی و پایانی کلمات با کرسی واو، دندانه، الف و بدون کرسی، صورت‌های مختلف نگارش الف مقصوره و مستوره در واژه‌ها، استفاده یا عدم استفاده از اعراب‌گذاری، و سایر علائم در مورد واژه‌هایی با شکل نوشتاری یکسان و تلفظ متفاوت، کاربرد یا حذف علائم همزه، تشدید، تنوین، مد و بعضی علائم مانند «ی» میانجی و ...هستند که سبب ایجاد مشکلاتی در فرایند جستجو و بازیابی اطلاعات شده‌اند. همچنین واژه‌های دو املایی (واژه‌هایی با واج یا آوای مشترک و شکل نوشتاری متفاوت)، استفاده از زبان محاوره و شکل عامیانه واژه‌ها در نوشتار، کسره اضافه و بدل‌های آن، جابه‌جایی ی و همزه در کلمات فارسی، نحوه نگارش «ه» غیرملفوظ و «ی» میانجی، وجود نقطه‌ها و دندانه‌های متعدد در بالا و پایین حروف نیز در این زمره‌اند (گل‌تاجی و بذرگر، ۱۳۸۹) مسائل یاد شده باعث می‌شود که کاربر در کاوش و بازیابی اطلاعات از

---

1. www.google.com, www.bing.com, www.yahoo.com

۲. در بعضی متون از بحث نوشتاری با عنوان نحوی نیز یاد می‌شود، اما در پژوهش حاضر براساس نظرات اساتید زبان و ادبیات فارسی، واژه نوشتاری استفاده شد که دربرگیرنده سایر مباحث نیز باشد.

3. Morphology & Semantic

موتورهای کاوش دچار سردرگمی شده و با نوشتن یک فرم از واژه، ناخواسته بسیاری نتایج را که حاوی صورت‌های دیگر یک واژه هستند از دست بدهد و فقدان توجه موتورهای کاوش به این ویژگی‌ها می‌تواند به این مسائل دامن بزند.

ب- ویژگی‌های معنایی: در مورد این ویژگی‌ها حسینی بهشتی (۱۳۸۶) معتقد است، دو مشکل عمده در ارتباطات معنایی واژگان وجود دارد که عبارت‌اند از:

- ✓ چندمعنایی، یعنی زمانی که یک کلمه واحد دارای معانی متعدد است.
- ✓ ترادف، به این معنی که کلمات متفاوت دارای یک معنی هستند.

هر دو پدیده مذکور، روابط شناسایی متداول در نظام بازیابی اطلاعات را مختل می‌سازند. به‌طور کلی از جمله ویژگی‌های معنایی زبان فارسی می‌توان همنامی یا واژه‌های یکسان با معانی متفاوت (واژگان مشترک لفظی)، چند معنایی، هم معنایی و ترادف را نام برد که همه این موارد در حوزه معنایی زبان می‌توانند همان چندگونگی‌های ذکر شده را ایجاد کنند، بدین معنی که کاربر در کاوش واژه با معنایی که در ذهن دارد دچار مشکل شده و گاهی موتورهای کاوش معانی دیگری غیر از آنچه که وی مدنظر دارد را بازیابی می‌کنند، مانند آنچه در جستجوی واژه «قلب» اتفاق می‌افتد. علاوه بر آنچه ذکر شد، «نبود استاندارد و شناور بودن ویژگی‌های رسم‌الخط و مفاهیم در زبان فارسی موجب گردیده که تقریباً به تعداد صفحات وب فارسی سبک و سیاق نگارشی برای این زبان به‌کار رفته باشد» (Shahidi & et al, 2005). بنابراین یافتن راهکارهایی جهت کاهش این مسائل مبتنی بر نتایج پژوهش‌ها در این حوزه امری ضروری به‌نظر می‌رسد.

پرسش‌های پژوهش

۱) مشکلات عمده مربوط به خط و زبان فارسی، براساس ویژگی‌های نگارشی (معنایی و نوشتاری) موجود در واژگان انتخابی در ارتباط با جستجو و بازیابی تصاویر در موتورهای کاوش موردمطالعه چه هستند؟

۲) کارآمدترین موتور کاوش در بازیابی تصاویر جهت انطباق با زبان فارسی کدام است؟

پیشینه پژوهش

بررسی‌های صورت گرفته در مورد پیشینه‌ها نشان می‌دهد که در اغلب موارد مسائل نوشتاری در ارتباط با بازیابی اطلاعات متنی از موتورهای کاوش برگزیده مدنظر بوده‌اند و در میان پیشینه‌های فارسی نیز پژوهشی که جنبه‌های نوشتاری و معنایی زبان را توأمان و در ارتباط با بازیابی تصاویر در نظر بگیرد انجام نشده است. در ادامه به برخی از پژوهش‌های صورت گرفته اشاره می‌شود.

به‌عنوان نمونه یافته‌های مربوط به پژوهش‌ها و مطالعات (عبداللهی و جوکار، ۱۳۸۸)، (گل‌تاجی و بذرگر، ۱۳۸۹) ، (آخشیک و فتاحی، ۱۳۹۱)، (ستوده و هنرجویان، ۱۳۹۱) ضمن تبیین و شناسایی بسیاری از ویژگی‌ها و مشکلات نگارشی زبان فارسی، در مجموع حاکی از آن است که رسم‌الخط فارسی یکی از متغیرهای عمده در ذخیره و بازیابی اطلاعات به زبان فارسی است و برخی ویژگی‌های املایی و ریخت‌شناسی زبان فارسی در فرایند کاوش مشکلاتی را پیش روی کاربران قرار می‌دهند به نحوی که عدم آگاهی و یا توجه کاربر به این ویژگی‌ها سبب ایجاد اختلال در کاوش و در نتیجه شکست جستجوی او می‌شود. از مسائل دیگر عدم توجه موتورهای کاوش وب، به شیوه‌های نگارش زبان فارسی به‌منظور بهبود عملکردشان در کاوش و به معنای اخص مواجهه با کاربران فارسی‌زبان است. حتی در مورد پایگاه‌های اطلاعاتی داخلی نیز بررسی‌ها نشان داد که چالش‌های ریختی شناخته شده زبان فارسی، تأثیر زیادی بر بازیابی اطلاعات در برخی از پایگاه‌های موردنظر دارند. همچنین مرور برخی پیشینه‌های خارج از کشور شامل پژوهش‌های لازارینیس[1] (۲۰۰۸، ۲۰۰۷)، ژانگ ولین[2] (۲۰۰۷)، لنداوسکی[3] (۲۰۰۸) نیز نشان می‌دهد ریخت‌شناسی کلمات و عبارات جستجو شده، به‌شدت بر بازیابی نتایج اثر دارد و موتورهای جستجو به‌جای تمرکز بر نیاز واقعی کاربران در جهت بهبود فرایند کاوش، بیشتر بر شکل کلیدواژه‌ها تکیه می‌کنند؛ حتی بعضی موتورهای جستجوی محلی نیز ریخت‌شناسی سؤالات را در نظر نمی‌گیرند و بنابراین جستجوی کاربر شکست می‌خورد. یافته‌ها همچنین حاکی از آن است که گوگل، در میان بسیاری از موتورهای جستجو مجهز به ویژگی پشتیبانی چندزبانه؛ به ترتیب بهترین جستجوگر با پشتیبانی چندزبانه است.

با در نظر گرفتن نزدیکی و شباهت‌های دو زبان فارسی و عربی خصوصا به لحاظ الفبایی، بررسی پیشینه‌های عربی موجود در این زمینه مانند مطالعات مقداد و لارج[4] (۲۰۰۱)، هامو[5] (۲۰۰۹)، و تاویلا و دیگران[6] (۲۰۱۰) در کل نشان‌دهنده این است که در رابطه با جستجو به زبان عربی در موتورهای کاوش؛ جستجوی واژه‌های عربی بدون پیشوند تعداد نتایج بازیابی شده را به‌طور چشمگیری کاهش می‌دهد. همچنین مشخص شد که گسترش سؤال با روش‌های مختلفی مانند اعراب‌گذاری، برای بهبود جستجو و بازیابی متون عربی ثمربخش است. دیگر اینکه موتور کاوش گوگل تقریباً در بیشتر موارد عملکرد بهتری نسبت به سایر موتورهای جستجوی مورد مطالعه داشته است.

1. Lazarinis
2. Zhang & Lin
3. Lewandowski
4. Moukdad & Large
5. Hammo
6. Tawileh & et al

بررسی پیشینه‌ها در مجموع حاکی از آن است که ویژگی‌ها و مشکلات زبانی به‌عنوان عامل مهمی در بحث جستجو و بازیابی اطلاعات مطرح هستند و در موارد بسیاری به آنها پرداخته شده است، اما در مورد زبان فارسی به دلیل ماهیت و خصوصیات منحصر به فرد آن، هنوز جای تحقیق و کار بسیار است. اضافه بر آن، پیشینه‌های داخلی یاد شده بر روی اطلاعات متنی تمرکز داشته‌اند و در مورد اطلاعات چند رسانه‌ای به نظر می‌رسد پژوهش پیش رو جزء اولین‌ها باشد.

روش‌شناسی پژوهش

پژوهش حاضر از نوع کاربردی است. برای پاسخ‌گویی به سؤالات پژوهش از روش ارزیابی با مشاهده مستقیم استفاده شد. بدین منظور، پس از بررسی منابع مرتبط و پیشینه‌های فارسی پژوهش با توجه به ویژگی‌ها و مشکلات ذکر شده در آنها برای زبان فارسی، نسبت به تهیه سیاهه محقق ساخته اقدام شد. در واقع کلیدواژه‌های موجود در سیاهه این پژوهش پیونددهنده میان ویژگی‌های نگارشی زبان فارسی و توانایی موتورهای کاوش در پاسخ‌گویی به این خصوصیات است. به این ترتیب که برای هر یک از ویژگی‌های نوشتاری و معنایی زبان فارسی واژه‌ای انتخاب شد تا به‌عنوان کلیدواژه کاوش، مبنا قرار گیرد به‌عنوان نمونه برای ویژگی «هم نامی» از واژه «شیر» برای آزمون موتورهای کاوش در رابطه با این ویژگی استفاده شد. سپس به جهت ماهیت بین‌رشته‌ای موضوع پژوهش، با استفاده از نظرات اساتید علم اطلاعات و دانش‌شناسی و زبان و ادبیات فارسی، از روایی سیاهه اطمینان حاصل شد (بدین معنی که آیا واژگان استفاده شده در سیاهه به‌درستی نمایانگر ویژگی مربوطه هستند؟). در نهایت سیاهه‌ای محقق ساخته شامل مشکلات و ویژگی‌های نوشتاری یعنی خصوصیاتی از زبان فارسی که با ریخت‌شناسی و ظاهر خط مرتبط هستند (پانزده ویژگی) و معنایی یعنی ویژگی‌های که مربوط به معناشناسی واژگان فارسی هستند (چهار ویژگی) که در مجموع در برگیرنده ۳۶ متغیر (کلمه) بود؛ تهیه شد[۱]. به‌منظور گردآوری داده‌ها، در تاریخ معین (شهریور ۱۳۹۲)، هریک از متغیرها توسط پژوهشگران به تفکیک، وارد بخش جستجوی تصاویر موتورهای کاوش مورد پژوهش شد و نتایج حاصل در جداولی ثبت شدند. در جداول یاد شده در بخش مسائل نوشتاری صورت‌های مختلف متصور برای هر واژه درج و سپس در میان پنجاه تصویر نخست بازیابی شده به‌صورت جداگانه تعداد یافته‌هایی که عیناً دارای همان کلیدواژه وارد شده توسط محققان،

---

۱. از میان ویژگی‌های نوشتاری و معنایی، گزینه‌هایی که قابلیت و بُعد بصری داشتند و مناسب به‌کارگیری در پژوهش حاضر بودند انتخاب شدند، به‌عنوان مثال استفاده از اسامی به‌جای افعال.

تعداد یافته‌هایی که حاوی دیگر صورت‌های نوشتاری درج شده در جدول و در نهایت یافته‌هایی با برچسبی متنی که حاوی هیچ‌یک از حالت‌های یاد شده نبودند (برای مثال نتایج بازیابی شده با برچسب متنی به زبان انگلیسی) شمارش و ثبت شدند. در بخش مسائل معنایی نیز به همین ترتیب عمل شد با این تفاوت که در مورد کلمات فاقد صورت‌های مختلف نو شتاری، مانند واژه‌های شیر و شور؛ در میان پنجاه نتیجه نخست بازیابی شده، تعداد تصاویری که حاوی معانی مختلف کلیدواژه جستجو شده بودند، شمارش و ثبت شدند. در این پژوهش تجزیه و تحلیل داده‌ها در دو سطح توصیفی و استنباطی انجام شد و برای این منظور از نرم‌افزار آماری SPSS و برای تفسیر از نرم‌افزار Word و Excel استفاده شده است.

یافته‌های پژوهش

پرسش اول: مشکلات عمده مربوط به خط و زبان فارسی، براساس ویژگی‌های نگارشی (نوشتاری و معنایی) موجود در واژگان انتخابی در ارتباط با جستجو و بازیابی تصاویر در موتورهای کاوش مورد مطالعه چه هستند؟

برای شناسایی مشکلات عمده نوشتاری و معنایی موجود در این واژگان، پس از وارد کردن کلیدواژه‌های موجود در سیاهه محقق ساخته در قسمت جستجوی تصاویر، نتایج بازیابی شده بررسی، شمارش، و ثبت شدند. پس از تجزیه و تحلیل آماری با استفاده از آزمون کای اسکور نتایج ذیل برای واژه‌ها (هر یک از واژه‌ها نمایانگر یکی از مشکلات نو شتاری و معنایی زبان فار سی هستند) به‌د ست آمده است.

جدول ۱. نتایج آزمون کای اسکور مربوط به حالات مختلف نگارشی کلیدواژه‌های مورد پژوهش در بخش نوشتاری

مسئله نوشتاری	واژه‌ها	مقدار	درجه آزادی	سطح معناداری	ضریب فی	نتیجه معنادار
ضبط واژگان لاتین	انفولانزا- آنفلوآنزا-آنفولانزا	۳/۶۷	۴	۰/۴۵۲	۰/۱۱۲	نیست
	تیتانیوم- تیتانیم	۰/۹۴۶	۲	۰/۶۲۳	۰/۰۶۸	نیست
واژگان دخیل و معادل آنها	کامپیوتر- رایانه	۱/۱۳	۲	۰/۵۶۷	۰/۰۶۹	نیست
	سیستم- نظام - سامانه	۱/۰۷۴	۴	۰/۸۹	۰/۰۵۴	نیست
واژگان مشتق	پستچی- پست چی	۷/۷۱۷	۲	۰/۰۲۱	۰/۲۲۱	است
واژگان مرکب	کتابخانه - کتاب‌خانه	۱/۱۹	۲	۰/۵۵۱	۰/۰۷۲	نیست
واژگان مشتق- مرکب	دانشسرا- دانش‌سرا	۳۹/۲۵	۲	۰/۰۰۰	۰/۵۴۷	است
	فناوری- فن‌آوری	۳/۹۶	۲	۰/۱۳۸	۰/۱۳	نیست

علائم جمع	گل‌ها- گل‌ها	۵/۴۱	۲	۰/۰۶۷	۰/۱۴	نیست
انواع جمع‌های فارسی و مکسر	مدارس- مدرسه‌ها	۷/۱۴	۲	۰/۰۲۸	۰/۲۰۶	است
	کتاب‌ها- کتب	۱۲/۱۱	۲	۰/۰۰۲	۰/۴۱	است
طریقه نگارش الف مقصوره	کسری- کسرا	۰/۰۶۹	۲	۰/۹۶	۰/۰۱۸	نیست
	مصلی - مصلا	۰/۰۰۱	۲	۰/۹۹۹	۰/۰۰۲	نیست
استفاده یا عدم کاربرد اعراب‌گذاری	مسکن- مُسکِن- مَسکَن	۰/۰۳۸	۲	۰/۹۹۹	۰/۰۸	نیست
	کره- کَره- کُره	۲/۷۲	۴	۰/۶۰۵	۰/۰۸۶	نیست
استفاده از تای منقوط	زکاة- زکات	۴/۱۱	۲	۰/۱۲۸	۰/۱۲۸	نیست
نحوه نگارش همزه میانی و پایانی کلمات	امضاء- امضا	۳/۰۳	۲	۰/۲۱۹	۰/۱۲	نیست
	شیء- شی	۹/۱	۲	۰/۰۱۱	۰/۲۶	است
	جبرئیل- جبرییل	۰/۹۳۴	۲	۰/۶۷	۰/۰۶	نیست
	مؤذن- مؤذن	۰/۰۵۷	۲	۰/۹۷	۰/۰۵۷	نیست
	مأمور- مأمور	۰/۶۰۳	۲	۰/۷۴	۰/۰۴۹	نیست
استفاده و عدم استفاده از تشدید	محمد - محمّد	۰/۱۲۲	۲	۰/۹۴۱	۰/۰۲۲	نیست
	زمرد - زمرّد	۱/۶۵	۲	۰/۴۳۱	۰/۰۹	نیست
کسره اضافه و بدل‌های آن	اعضا بدن- اعضای بدن- اعضاء بدن	۰/۶۴	۴	۰/۹۵۸	۰/۰۴۷	نیست
واژه‌های دو املائی	آذوقه- آزوقه	۱/۹۸	۲	۰/۳۲۷	۰/۱۳۳	نیست
	تهران- طهران	۰/۹۱۸	۲	۰/۶۳	۰/۰۶	نیست
جابجایی ی و همزه در کلمات فارسی	پاییز- پائیز	۳/۵۱	۲	۰/۱۷	۰/۱۱۷	نیست
	آیینه- آئینه	۱/۳۳	۲	۰/۵۱	۰/۱۰۲	نیست
نحوه نگارش ه غیرملفوظ و ی میانجی	جامعه‌اطلاعاتی، جامعه‌ی اطلاعاتی، جامعهٔ اطلاعاتی	۱/۰۱۲	۴	۰/۹۰۸	۰/۰۵۳	نیست
استفاده از زبان محاوره	خانه - خونه	۸/۳۲	۲	۰/۰۱۶	۰/۲۰۷	است
کاربرد و حذف مد در کلمات فارسی	پیشاهنگ- پیشآهنگ	۰/۴۵۸	۲	۰/۷۹۵	۰/۰۴۵	نیست

در جدول شماره ۱، مقدار کای اسکور، درجه آزادی، سطح معنادار و ضریب فی برای تک‌تک واژه‌ها در بخش نوشتاری محاسبه شده است. همان‌طور که در جدول یاد شده مشهود است، در رابطه با واژه‌هایی که سطح معنادار بزرگ‌تر از آلفا ۰/۰۵ دارند با اطمینان ۰/۹۵؛ رابطه معنی‌داری بین ویژگی‌های نوشتاری و مشکلات مربوط به بازیابی تصاویر از موتورهای کاوش مورد مطالعه وجود ندارد؛ اما سطح

معنادار به‌دست آمده برای واژه‌های پستچی- پست‌چی (واژگان مشتق)، دانشسرا- دانش‌سرا (واژگان مشتق- مرکب)، مدارس- مدرسه‌ها و کتاب‌ها- کتب (انواع جمع‌های فارسی و مکسر عربی)، شیء- شی (نگارش همزه پایانی بدون کرسی) و خانه- خونه (استفاده از زبان محاوره) از آلفا مفروض ۰/۰۵ کوچک‌تر است، پس با اطمینان ۰/۹۵ رابطه معنی‌داری بین این ویژگی‌ها، و مشکلات مربوط به جستجو بازیابی تصاویر از موتورهای کاوش مورد مطالعه وجود دارد و در مجموع، مسائل مربوط به پیوسته‌نویسی و جدانویسی کلمات مشتق، مشتق- مرکب، انواع جمع‌های فارسی و مکسر عربی، نگارش همزه بدون کرسی و استفاده از زبان محاوره؛ جزء مشکلات عمده مربوط به جستجو و بازیابی تصاویر از موتورهای کاوش مورد پژوهش هستند.

جدول ۲. نتایج آزمون کای اسکور مربوط به معانی گوناگون کلیدواژه‌های مورد پژوهش در بخش معنایی

مسئله معنایی	واژه‌ها	مقدار	درجه آزادی	سطح معناداری	ضریب فی	نتیجه معنادار
همنامی یا واژه‌های یکسان با معانی متفاوت	شیر (حیوان)- شیر (لبنی)- شیر (شیر آب)	۱/۶۳	۲	۰/۴۴	۰/۱۰۴	نیست
واژگان هم نویسه با معانی متفاوت	شور (طعم)- شور (اشتیاق)	۱/۰۴	۲	۰/۵۹	۰/۱۲۴	نیست
چند معنایی	قلب (عضو بدن)- قلب (خاطر و ضمیر)- قلب (وارونه کردن)- قلب (مرکز)	۶/۳۷	۲	۰/۰۴۱	۰/۲۰۹	است
	نوک (اشیاء)- نوک (پرنده)	۰/۵۹	۲	۰/۷۴۵	۰/۰۹	نیست
هم‌معنایی و ترادف	دریای خزر- دریای کاسپین- دریای مازندران	۲/۱۶۸	۴	۰/۷۰۵	۰/۰۷	نیست

همان‌طور که در جدول ۲ مشهود است، در رابطه با واژه‌هایی که سطح معنادار بزرگ‌تر از آلفا ۰/۰۵ دارند با اطمینان ۰/۹۵؛ رابطه معنی‌داری میان ویژگی‌های معنایی و مشکلات مربوط به بازیابی تصاویر از موتورهای کاوش مورد مطالعه وجود ندارد؛ اما سطح معنادار به‌دست آمده برای معانی مختلف واژه قلب (چند معنایی) از آلفا مفروض ۰/۰۵ کوچک‌تر است، پس با اطمینان ۰/۹۵ رابطه معنی‌داری بین این ویژگی معنایی و مشکلات مربوط به جستجو بازیابی تصاویر از موتورهای کاوش مورد مطالعه وجود دارد و در مجموع مبحث معنایی، مسئله مربوط به چند معنایی، از مشکلات عمده مربوط به جستجو و بازیابی تصاویر از موتورهای کاوش مورد پژوهش است.

پرسش دوم: کارآمدترین موتور کاوش در بازیابی تصاویر جهت انطباق با زبان فارسی کدام است؟

برای شناسایی کارآمدترین موتور کاوش جهت انطباق و توجه به ویژگی‌های نوشتاری و معنایی زبان فارسی، از میان سه موتور جستجوی مورد پژوهش؛ پس از وارد کردن کلیدواژه‌های موجود در سیاهه محقق ساخته در قسمت جستجوی تصاویر، نتایج بازیابی شده بررسی و ثبت شدند. پس از تجزیه و تحلیل آماری با استفاده از آزمون فریدمن نتایج ذیل به‌دست آمده است.

جدول ۳. نتایج آزمون فریدمن مربوط به میانگین موتورهای کاوش، در بازیابی حالات مختلف نگارشی کلیدواژه‌های مورد پژوهش در بخش نوشتاری

مسئله نوشتاری	واژه‌ها	بینگ	یاهو	گوگل	قوی‌ترین موتور کاوش
ضبط واژگان لاتین	آنفولانزا- انفولانزا – آنفلوآنزا	۲۵/۰۲	۳۱/۸۷	۳۱/۴۷	یاهو
	تیتانیوم- تیتانیم	۳۲/۱۲	۳۶/۴	۳۳/۲۵	یاهو
واژگان دخیل و معادل آنها	کامپیوتر- رایانه	۳۳/۵۳	۴۲/۴۸	۴۴/۴۳	گوگل
	سیستم- نظام- سامانه	۳۶/۹۶	۴۷/۷۹	۴۱/۰۳	یاهو
واژگان مشتق	پستچی- پست‌چی	۶۴/۳۲	۲۴/۷۶	۳۴/۸۲	گوگل
واژگان مرکب	کتابخانه- کتاب‌خانه	۳۶/۳۵	۳۷/۸۱	۴۱/۸۸	گوگل
واژگان مشتق- مرکب	دانشسرا- دانش‌سرا	۳۴/۱۳	۳۴/۸۲	۱۴/۲۸	یاهو
	فناوری- فن‌آوری	۴۰/۲۶	۴۳	۳۳/۵۹	یاهو
علائم جمع	گل‌ها- گل‌ها	۴۲/۰۲	۴۲/۷۳	۴۳/۲۲	گوگل
انواع جمع‌های فارسی و مکسر	مدارس- مدرسه‌ها	۲۵/۸۷	۳۶/۷	۳۹/۵۸	گوگل
	کتاب‌ها- کتب	۱۳/۳	۱۱/۶۵	۱۴/۲۶	گوگل
طریقه نگارش الف مقصوره	کسری- کسرا	۳۵/۸۴	۳۱/۳۹	۴۰/۸۸	گوگل
	مصلی- مصلا	۴۴/۷۷	۴۲/۷۷	۳۲/۶۹	بینگ
استفاده یا عدم کاربرد اعراب‌گذاری	مسکن- مُسَکِن- مَسکَن	۳۳/۰۶	۴۲/۴۴	۴۸/۶۴	گوگل
	کره- کَره- کُره	۳۳/۰۶	۴۲/۴۴	۴۸/۶۴	گوگل
استفاده از تای منقوط	زکاۀ- زکات	۴۱/۳۹	۴۴/۰۳	۳۹/۸۶	یاهو
نحوه نگارش همزه میانی و پایانی کلمات	امضاء- امضا	۳۴/۵۶	۳۵/۶	۳۵/۶۹	گوگل
	شیء- شی	۳۵/۱۷	۳۵/۳۶	۳۵/۴۳	گوگل
	جبرئیل - جبرییل	۴۰/۸	۳۹/۶	۴۱/۳۲	گوگل
	مؤذن- مؤذن	۴۱/۵۵	۴۱	۳۹/۰۳	بینگ

	مأمور– مأمور	۴۱/۲۱	۴۰/۰۷	۴۴/۸۵	گوگل
استفاده و عدم استفاده از تشدید	محمد – محمّد	۴۵/۰۲۲	۴۴/۹۷	۴۰/۹۷	بینگ
	زمرد – زمرّد	۲۶/۰۳	۴۳/۳	۲۹/۰۳	یاهو
کسره اضافه و بدل‌های آن	اعضا بدن– اعضای بدن– اعضاء بدن	۳۹/۹۵	۳۷/۹۲	۲۷/۷۷	بینگ
واژه‌های دو املائی	آذوقه– آزوقه	۳۴/۱۶	۳۰/۶۲	۲۵/۷۴	بینگ
	تهران– طهران	۳۱/۸۸	۳۶/۱۳	۴۰/۶۷	گوگل
جابجایی ی و همزه در کلمات فارسی	پاییز– پائیز	۴۳/۱۶	۴۵/۹۱	۴۰/۱۶	یاهو
	آیینه– آئینه	۲۸/۴۵	۲۶/۸	۳۱/۹۷	گوگل
نحوه نگارش ه غیرملفوظ و ی میانجی	جامعه‌اطلاعاتی، جامعه‌ی اطلاعاتی، جامعهٔ اطلاعاتی	۴۹/۲۳	۴۲/۲۹	۳۵/۲۳	بینگ
استفاده از زبان محاوره	خانه– خونه	۲۹/۸۹	۳۸/۷۱	۴۴/۰۸	گوگل
کاربرد و حذف مد در کلمات فارسی	پیشاهنگ– پیشآهنگ	۴۴	۴۰	۲۸	بینگ

با استفاده از آزمون فریدمن، میانگین واژه‌ها، به تفکیک موتورهای کاوش مورد پژوهش محاسبه شد. نتایج جدول شماره ۳ نشان می‌دهد که در بخش نوشتاری به ترتیب، موتورهای کاوش گوگل، یاهو و در آخر بینگ بیشترین تطبیق را با ویژگی‌های نوشتاری زبان فارسی دارند.

جدول ۴. نتایج آزمون فریدمن مربوط به میانگین موتورهای کاوش، در بازیابی معانی گوناگون کلیدواژه‌های مورد پژوهش در بخش معنایی

مسئله معنایی	واژه‌ها	بینگ	یاهو	گوگل	قوی‌ترین موتور کاوش
واژگان هم نویسه با معانی متفاوت	شور (طعم)– شور (اشتیاق)	۱۶/۸۱	۱۷/۴	۱۹	گوگل
چند معنایی	نوک (اشیاء)– نوک (پرنده)	۶/۸	۶/۷۴	۹/۲۷	گوگل
	قلب (عضو بدن)–قلب (خاطر و ضمیر)	۴۲/۳۷	۴۲/۳۷	۳۸/۱۲	یاهو و بینگ
هم‌معنایی و ترادف	دریای خزر– دریای کاسپین– دریای مازندران	۴۵/۸۴	۴۵/۸۱	۳۳/۱۱	بینگ
هم‌نامی	شیر (حیوان)– شیر (لبنی)	۴۴/۳۳	۴۵/۱۸	۴۶/۱۲	گوگل

همان‌طور که در جدول شماره ۴ مشاهده می‌شود، محاسبه میانگین واژه‌ها با استفاده از آزمون فریدمن نشان می‌دهد که در بخش معنایی نیز، موتورهای کاوش گوگل، بینگ و در آخر یاهو به ترتیب بیشترین انطباق را با ویژگی‌های معنایی زبان فارسی دارند. در مجموع نتایج جداول ۳ و ۴ نشان می‌دهد که در دو بخش ویژگی‌های نوشتاری و معنایی، از بین سه موتور کاوش مورد پژوهش؛ موتور جستجوی گوگل، بیشترین انطباق را با این ویژگی‌ها دارد، و پس از آن موتور کاوش بینگ و یاهو در جایگاهی برابر قرار می‌گیرند.

نتیجه

با توجه به یافته‌های پژوهش حاضر، بین ویژگی‌ها و مسائل مربوط به پیوسته‌نویسی و جدانویسی واژگان مشتق، مشتق- مرکب، انواع جمع‌های فارسی و مکسر عربی، نگارش همزه بدون کرسی و استفاده از زبان محاوره در بخش نوشتاری و همچنین مشکل مربوط به چند معنایی در بخش معنایی و مشکلات مربوط به بازیابی تصاویر از موتورهای کاوش مورد مطالعه رابطه معنی‌داری وجود دارد. به‌عنوان نمونه با انتخاب کلیدواژه پستچی توسط کاربر و بازیابی تصاویر مربوط به این واژه، کاربر بدون اینکه آگاه باشد، در بسیاری موارد از دستیابی به تصاویر با برچسب پست‌چی محروم مانده و تعداد قابل‌توجهی از نتایج را از دست می‌دهد. همچنین با دقت در یافته‌ها می‌توان دریافت که واژگان عربی موجود در زبان فارسی مانند آنچه در بحث انواع جمع‌ها و نگارش همزه آمد، سهم قابل‌توجهی از این مشکلات را به خود اختصاص داده‌اند، در بخش ویژگی‌های معنایی نیز، مسئله چندمعنایی، مشکل عمده‌ای محسوب می‌شود، نظیر آنچه کاربر در کاوش واژه قلب با آن روبرو می‌شود بدین معنا که در بسیاری موارد کاربر تنها تصاویری را که حاوی یکی از معانی این واژه هستند در میان نتایج ابتدایی بازیابی می‌کند و در نتیجه دچار سردرگمی می‌شود.

در نتیجه مسائل یاد شده، از مشکلات عمده زبان فارسی در جستجو و بازیابی تصاویر از موتورهای کاوش مورد پژوهش به‌شمار می‌آیند، که با توجه به ویژگی‌های این موتورهای کاوش، قابل تعمیم به انواع دیگر موتورهای جستجوگر و نیز جستجوی سایر گونه‌های اطلاعات نیز هستند. یافته‌های مذکور با نتایج حاصل از پژوهش عبداللهی و جوکار (۱۳۸۸) مبنی بر اینکه هیچ‌کدام از موتورهای کاوش، چالش‌های شیوه‌های نگارش فارسی را به‌منظور بهبود نتیجه کاوش، مورد توجه قرار نداده‌اند؛ هم‌سویی دارد. همچنین بررسی یافته‌ها، هم‌سویی آنها با نتایج حاصل از پژوهش لازارینیس (۲۰۰۷) مبنی بر اینکه موتورهای کاوش نتایج مختلفی را برای واژه‌های متفاوت از نظر ریخت‌شناسی و نوشتاری بازیابی می‌کنند؛ نشان می‌دهد. نتایج

پژوهش راثی ساربانقلی (۱۳۸۵) هم حاکی از آن است که ویژگی‌های املایی و ریخت‌شناسی زبان فارسی در کاوش اطلاعات از موتورهای کاوش، مشکلاتی ایجاد می‌کند که با یافته‌های پژوهش حاضر هم‌سویی دارد.

همچنین نتایج دو بخش نوشتاری و معنایی مبنی بر برتری قابلیت‌های موتور کاوش گوگل نسبت به یاهو و بینگ، با نتایج حاصل از پژوهش ژانگ و لین (۲۰۰۷) که گوگل را بهترین موتور با ویژگی پشتیبانی چندزبانه در بین موتورهای کاوش مورد بررسی‌شان می‌دانند، هم‌سویی دارد؛ اما در مجموع، موتورهای کاوش مورد پژوهش نسبت به ویژگی‌های نوشتاری و معنایی زبان فارسی توجه کافی ندارند، و بسیاری از ویژگی‌های آن را در هنگام جستجو و بازیابی تصاویر نادیده می‌گیرند. این مسئله موجب می‌شود که احتیاج فارسی‌زبانان به موتورهای کاوش بومی که مبتنی بر ویژگی‌های زبانی خودشان طراحی شده باشد بیش از پیش احساس شود. از سوی دیگر موتورهای جستجوی گوگل، بینگ و یاهو که جزء موتورهای کاوش محبوب هستند نیز باید نسبت به برآورده ساختن نیازهای کاربران غیرانگلیسی زبانشان بیشتر تلاش کنند. همان‌طور که نتایج این پژوهش نشان داد، نادیده انگاشتن و یا کم توجهی به شاخصه‌ها و ویژگی‌های زبانی کاربران موجب بروز مسائلی در امر جستجو و بازیابی اطلاعات می‌شود که در نهایت از دست رفتن اطلاعات مفید و یا بازیابی اطلاعات ناخواسته را به همراه خواهد داشت. در ادامه با توجه به یافته‌های پژوهش، پیشنهادهایی به شرح زیر ارائه می‌شود:

- ✓ با توجه به یافته‌های پژوهش، بسیاری از مشکلات نگارشی فرایند کاوش و بازیابی تصاویر از موتورهای کاوش مربوط به گوناگونی نوشتار یک مفهوم واحد هستند، لذا تلاش برای یکپارچگی شیوه‌های نگارشی در محیط وب، به‌عنوان نمونه تدوین شیوه‌نامه‌ای استاندارد برای نوشتن در وب می‌تواند تا حدی از این مسائل بکاهند.
- ✓ به‌منظور آگاهی کاربران فارسی‌زبان از گوناگونی‌های نوشتاری یاد شده، تدوین شیوه‌نامه‌ای آموزشی برای کاربران، با تأکید بر مشکلات عمده شناسایی شده در این پژوهش، می‌تواند کاربران را در بازیابی موفق‌تر یاری کند.
- ✓ ایجاد سازوکاری در موتورهای کاوش جهت آگاهی دادن به کاربران، همزمان با درج کلیدواژه در هنگام جستجو، از طریق نمایش مترادفات، معادل فارسی واژگان دخیل و همچنین صورت‌های املایی گوناگون می‌تواند از سردرگمی کاربر پیشگیری کند.

- ✓ در انتها پیشنهاد می شود متولیان امر، نظیر فرهنگستان زبان و ادب فار سی نظارت بیشتری نسبت به رعایت یکپارچگی و هماهنگی در متون تولید شـده به‌صـورت چاپی و الکترونیک داشته باشـند تا ضــمن حراســت از زبان غنی فارســی، از بروز برخی چالش‌هایی از قبیل آنچه در این پژوهش شناسایی و بررسی شد جلوگیری نمایند.

نظر به نتایج پژوهش پیش رو، در بســیاری موارد عدم توجه به ویژگی‌ها و ظرایف نوشـتاری و معنایی زبان فار سی موجب بروز ا شکالات جدی در فرایند جستجو و بازیابی تصاویر می شود؛ از همین رو در ادامه پیشـنهادهای پژوهشـی که انجام آن در آینده می‌تواند زمینه‌سـاز حل این چالش‌ها باشـد ارائه می‌شود:

- ✓ انجام پژوهش‌هایی بین‌رشـته‌ای با رویکرد سـببشـناسـی و چرایی مسـائل به‌وجود آمده در اثر رویارویی ویژگی‌های خط و زبان فارسی با فناوری‌های نو.
- ✓ انجام پژوهش‌هایی با هدف سنجش میزان آشنایی و توجه کاربران فارسی‌زبان موتورهای کاوش به ویژگی‌های خط و زبان فارسی و ارزیابی تأثیر آن بر موفقیت جستجو.
- ✓ بررسی مقایسه‌ای بین موتورهای کاوش بومی با انواع غیربومی آن با معیار توجه به ویژگی‌های خط و زبان فارسی.
- ✓ برر سی فرایند نمایه سازی تصاویر در موتورهای کاوش پرکاربرد و شنا سایی نقاط ضعف و قوت آنها با توجه به شاخصه‌های نوشتاری و معنایی زبان فارسی.
- ✓ بررسی امکان ایجاد راهبردهای فرا کاوش جهت بهبود جستجو و بازیابی تصاویر به زبان فارسی.

کتابنامه

اسلامی، محرم (۱۳۸۱). «دشواری‌های پردازش رایانه‌ای خط فارسی». *فصلنامه نشر دانش*، ۱۷(۳)، ۳۲-۲۸.

اکبری نژاد، سعید (۱۳۷۶). «فاصله خالی میان واژه‌ها در ذخیره و بازیابی رایانه‌ای اطلاعات». *فصلنامه کتاب*، ۸(۱)،۴۹- ۵۶.

آخیشک، سمیه سادات؛ فتاحی، رحمت‌الله (۱۳۹۱). «تحلیل چالش‌های پیوسته‌نویسی و جدانویسی واژگان فارسی در ذخیره و بازیابی اطلاعات در پایگاه‌های اطلاعاتی». *فصلنامه کتابداری و اطلاع‌رسانی*، دوره شانزدهم، شماره سوم، ۹- ۳۰.

حری، عباس (۱۳۷۲) «کامپیوتر و رسم‌الخط فارسی».*فصلنامه تحقیقات اطلاع‌رسانی و کتابخانه‌های عمومی*، ۳(۱)، ۶- ۱۱.

حسینی بهشتی، ملوک‌السادات (۱۳۸۶). «معنی‌شناسی واژگانی فرااصطلاحنامه و بازیابی اطلاعات». کتاب ماه کلیات مجموعه اطلاع‌رسانی و کتابداری، ۱۰ (۱۰) ، ۳۰- ۳۷.

داورپناه، محمدرضا (۱۳۸۷). *جستجوی اطلاعات علمی و پژوهشی در منابع چاپی و الکترونیکی*. تهران: چاپار؛ دبیزش.

راثی ساربانقلی، محمد (۱۳۸۵). «مشکلات جستجو و بازیابی اطلاعات به زبان فارسی در اینترنت، مطالعهٔ موردی: کاربران مرکز اینترنت دانشگاه اسلامی واحد شبستر». *فصلنامه کتاب، کتابداری و اطلاع‌رسانی*، ۱۷(۳)، ۱۷۹- ۱۹۶.

ستوده، هاجر؛ هنرجویان، زهره (۱۳۹۱). «مروری بر دشواری‌های زبان فارسی در محیط دیجیتال و تأثیرات آنها بر اثربخشی پردازش خودکار متن و بازیابی اطلاعات»، *کتابداری و اطلاع‌رسانی*، ۱۵(۴)، ۹۲-۵۹.

عبدالهی نورعلی، محمدصادق؛ جوکار، عبدالرسول (۱۳۸۸). « چالش‌های شیوه نگارش زبان فارسی در بازیابی اطلاعات از موتورهای کاوش وب». *مطالعات تربیتی و روانشناسی*، ۱۰(۲) ، ۶۷- ۹۰.

گل‌تاجی، مریم؛ بذرگر، سعیده (۱۳۸۹). «بررسی مشکلات ریخت‌شناسی زبان فارسی در سه پایگاه اطلاعاتی مرکز منطقه‌ای اطلاع‌رسانی علوم و فناوری، پژوهشگاه اطلاعات و مدارک علمی ایران و جهاد دانشگاهی»، *کتابداری و اطلاع‌رسانی*، ۱۳(۲)،۲۲۲-۱۹۹.

مرتضائی، لیلا (۱۳۸۰). «مسائل زبان و خط فارسی در ذخیره‌سازی و بازیابی اطلاعات». *علوم و فناوری اطلاعات*، ۱۷(۱) ، ۲۴- ۲۹.

نوتس، گری (۱۳۸۲). «راهبردها و شیوه‌های جستجو در اینترنت»، ترجمه سیمین نیازی، *فصلنامه کتاب، کتابداری و اطلاع‌رسانی*، ۱۴(۳)، ۱۲۶-۱۳۳.

نوروزی، علیرضا؛ ولایتی، خالید (۱۳۸۹). *نمایه‌سازی موضوعی: نمایه‌سازی مفهومی*. تهران: چاپار.

Hammo, B. (2009). Towards enhancing retrieval effectiveness of search engines for diacritisized Arabic documents. Information Retrieval, 12(3), 300-323. Available at: http://link.springer.com/article/10.1007/s10791-008-9081-9 Accessed (2012 July 19).

Internet world stats: Usage and population statistics, 2015 May 31. http://www.internetworldstats.com/stats7.htm, Accessed (2014 March 25).

Lazarinis, F. (2007, C). At the sharp END evaluating the searching capabilities of commerce websites in a non-English language A Greek case study. Online Information Review, 31(6), 881-891. Available at: http://www.emeraldinsight.com/journals.htm?articleid=1640585. Accessed (2012 July 17).

Lazarinis, F. (2007, A). Web retrieval systems and the Greek language: do they have an understanding? Journal of information science, 33(5), 622-636.

Lazarinis, F. (2008, B). Improving concept-based web image retrieval by mixing semantically similar Greek queries. Program: electronic library and information systems, 42(1), 56-67. Available at:

http://www.emeraldinsight.com/journals.htm?articleid=1674242. Accessed (2012 July 17).

Lewandowski, D. (2008). Problems with the use of Web search engines to find results in foreign languages. Online Information Review, 32(4), 668 – 672. Available at: http://www.emeraldinsight.com/journals.htm?articleid=1747662 . Accessed (2012 June 15).

Liu,Y. & et al. (2007). A survey of content-based image retrieval with high-level semantics. Pattern Recognition, 40(1), 262–282. Available at: http://www.sciencedirect.com/science/article/pii/S0031320306002184. Accessed (2012 January 1).

Moukdad, H., Large, A. (2001). Information Retrieval from Full-Text Arabic Databases: Can Search Engines Designed for English Do the Job? Libri, 51(3), 63- 74. Available at: www.**libri**journal.org/pdf/2001-2pp63-74.pdf. Accessed (2014 May 4)

Notess, G.R. (1997). Internet Search Techniques and Strategies. Online, 21(4), 63-66. Available at: http://www.questia.com/library/1G1-19545628/internet-search-techniques-and-strategies. Accessed (2013 December 11).

TASI (Technical Advisory Service for Images), 2008 January 25, Review of Image Search Engines, Accessed 2013 April 20, available http://www.jiscdigitalmedia.ac.uk/guide/review-of-image-search- engines/

Tawileh, W., Mandl, Th. and Griesbaum, J. (2010). Evaluation of five web search engines in Arabic language. 10th International Conference on Intelligent Systems Design and Applications, 2010, Cairo, Egypt, retrieved 2013 August 6, available at: www.kde.cs.uni-kassel.de/conf/lwa10/.../ir1.pdf.

Zachary, J., Lyengar, s. s. and Barhen, J. (2001). Content based image Accessed and information theory: a general approach. Journal of the American society for information science and technology, 52(10), 840-852. Available at: http://onlinelibrary.wiley.com/doi/10.1002/asi.1138/pdf. Retrieved 2013 January 02.

Zhang, J., Lin, S. (2007). Multiple language supports in search engines. Online Information Review, 31(4),516-532. Available at: http://www.emeraldinsight.com/journals.htm?articleid=1621798 . Accessed (2012 July 13).